Audio in VR
- Audio is very important in VR
- High Quality audio provides:
  - Increased realism
  - Strong immersive sense
  - Strong positional cues
  - Extra information about the environment
  - The shape of the world
- What does 'High Quality' mean?

VR sound environment
- VR equipment is a really bad place to be trying to create sound.
- CAVE
  - Stand in a glass box
  - Pretend you can't hear the echoes.
  - Also hard to place speakers

VR sound environment (2)
- Semi-immersive VR theatre
  - Sit in a big cylinder
  - Reflects sound in a very strange way.

Workbench
- Not so bad but still...
- Big flat screen 1 metre in front of you
- Sound coming from surround speakers
- Creates echoes inappropriate for scene

Audio equipment in VR
- So, most of VR audio is based on very high quality headphones
- Use head tracking to get position and orientation
- Generate sounds as if from sources
- Play to user
- Problem solved!
- ...sadly, not.
**A world of sound**

- You are surrounded by sound all the time - Silence is unheard of!
- The environment affects (shapes?) the sound you hear
  - Size
  - Shape
  - Materials

**What do you hear?**

![Time Amplitude Graph]

**Rendering sound - ‘Auralization’**

- To generate correct echoes must model sound behaviour in the space
- Rooms are complex
- Filled with different materials
  - Reflective
  - Absorbant
  - Frequency filtering
- Just like rendering light

**Putting sound in VR environments**

- What sound?
- ‘Ambient’ sounds
  - ‘Surround’ sound
  - Often use recorded sounds
- Positional sounds
  - Designed to give a strong sense of something happening in a particular place
  - Also often provided by using recorded sounds

**Positional sound**

- Using sound to create the sense of active things in the environment
  - Enhances presence
  - Enhances immersion
- Need to deal with many components
  - Reflections (echoes)
  - Diffraction effects

**Consider VisClim**

- Scene in Linköping’s Storatorget
- Surrounding environment
  - Vehicles? - Several roads nearby
  - People? - Many people in the square
- Weather noise effects
  - Rainfall
  - Snowfall (no sound but damping effect)
Air traffic control

- Simulation so
  - No ‘ambient’ sound required
  - No aircraft noises
  - No realism wanted?
- Positional warnings?
  - Designed to draw the users attention to the location of a problem
  - Which may be out of the field of view

Creating positional sound

- Amplitude
  - stereo (or more)
- Synchronisation
  - Audio delays
- Frequency
  - HRTF

Positional sound - amplitude

- Generate audio from position sources
- Calculate amplitude from distance
- Include damping factors
  - Air conditions
  - Snow
  - Directional effect of the ears

Positional sound - Synchronisation

- Ears are very precise instruments
- Very good at hearing when something happens after something else
- Use this to help define direction
  - Difference in amplitude gives only very approximate direction information
Positional sound - synchronisation

- 30 centimetres
- =0.001 seconds!
- Human can hear ≤ 700μS

3D positional sound

- Humans have stereo ears
  - Two sound pulse impacts
  - One difference in amplitude
  - One difference in time of arrival
- How is it that a human can resolve sound in 3D?
- Should only be possible in 2D?

Positional sound - Frequency

- Frequency responses of the ears change in different directions!
- You hear a different frequency filtering in each ear
- Use that data to work out 3D position information
- How do we mimic that in headphones?

Head-Related Transfer Function

- Define a frequency transfer function
  - for each ear
  - for every direction around the head
- HRTF varies from person to person
  - Must measure each individual user!
  - Measurements involve microphones inserted deep in the user's ears.
  - Use movable sound source to measure response

Using HRTF's

- HRTF's are 3D
- Depend on ear shape (Pinna) and resonant qualities of the head!
- Allows positional sound to be 3D
- Computationally difficult
  - Originally done in special hardware (Convolotron)
  - Now can be done in real-time using DSP
### Audio rendering summary
- Rendering audio is really, really hard
- Much bigger problem than lighting
- Material properties are more complex
  - Can’t fake it as easily
  - Properties are always a problem
- Many good methods exist but the problem is too computationally hard for general use at present

### What can we do today?
- Constraint is real time audio rendering
- Simple (reflectionless) stereo positional sound
  - Using amplitude
  - Using synchronization
  - Using HRTF frequency filtering
- Useful for audio cues and simple environmental sounds

### What can’t we do today
- Full interactive audio rendering
- Still too large a computational problem
  - Reflections and diffractions too complex
  - For real situations at least
- Simple scenes do provide some scope
  - Buildings – hard materials, flat planes
  - Could handle VisClim in real time?

### Voice interaction
- Voice input for control
  - Continuous?
  - Discrete?
- Voice output for information
  - Positional? – Alerts
  - Arbitrary? – Purely informational

### Voice output
- Speech synthesis is quite sophisticated
- Can create a reasonably human-sounding synthetic voice
  - Quite understandable
  - Often not too pleasant to listen to
- Can use sampled sound and signal processing to create a real sounding voice
  - In more limited circumstances

### Voice output
- Can combine voice synthesis with positional sound to create a warning
  - Use position as an ‘attractor’
  - Draw user’s attention to the location of the problem
- (Maybe?) no need to worry about complex sound rendering
Voice recognition

- Potentially very powerful interaction mode
- Can free the user
  - From keyboard
  - From 2D or 3D mouse
- Been in development for 20 years
  - with limited success

Voice recognition: The dream

Voice recognition: The reality

Continuous speech recognition

- Requires
  - Very good sound equipment
    - High quality microphone
    - High quality sound hardware on computer
  - Minimum of 500MHz processing power
    - Probably much more
  - Lots and lots of physical memory
    - database of grammar and word structures

Discrete speech pattern recognition

- Create a set of command phrases
  - As small as possible
  - As distinct as possible ("on"/"off" are bad!)
- Use audio pattern matching to determine if one of the command phrases has been said.
- Have to get it’s attention somehow
  - Like a button?

Voice recognition in ATC

- Air traffic control application
  - Has discrete set of commands
  - Often want to turn features on and off
    - "enable" and "disable"
  - Can build a command set
- Use button to get its attention
- Error rate <5% (on a good day)
ATC command set

- restart scenario: restart scenario
  Say this part
  This is sent to the application
- Currently have about 110 commands
  - 25 of which are flight selection

ATC command set(2)

- height scale 1: scalefactor 1
- height scale 3: scalefactor 3
- time factor 1: updatefreq 1
- time factor 4: updatefreq 4
- trajectory inactive disable: trajmode 0 0
- trajectory inactive enable: trajmode 0 1
- focus on flight S K 2 3 1: airplane focus on SK231
- focus on flight S K 2 3 2: airplane focus on SK232
- focus on flight S K 2 3 4: airplane focus on SK234

Summary: Voice interaction

- Speech synthesis works quite well
  - Allows for easy informative feedback
- Speech recognition is harder
  - General case still impossible
- Defined command set can be worked with
  - Recognition rates can be very high
  - Still not perfect